

#3 4-3-02  
IDS w/Refer  
Docket No.: M2082.0000/P000  
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:  
Satoshi Hasegawa

Application No.: Not Yet Assigned

Group Art Unit: N/A

Filed: Herewith

Examiner: Not Yet Assigned

For: DEVICE AND METHOD FOR  
PROCESSING AUDIO INFORMATION



CLAIM FOR PRIORITY AND SUBMISSION OF DOCUMENTS

Commissioner for Patents  
Washington, DC 20231

Dear Sir:

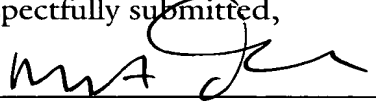
Applicant hereby claims priority under 35 U.S.C. 119 based on the following prior foreign application filed in the following foreign country on the date indicated:

<u>Country</u>	<u>Application No.</u>	<u>Date</u>
Japan	2001-009564	January 17, 2001

In support of this claim, a certified copy of the said original foreign application is filed herewith.

Dated: January 17, 2002

Respectfully submitted,

By 

Mark J. Thronson

Registration No.: 33,082

DICKSTEIN SHAPIRO MORIN &  
OSHINSKY LLP

2101 L Street NW

Washington, DC 20037-1526

(202) 785-9700

Attorneys for Applicant

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

J1002 U.S. PTO  
10/046719  
01/17/02

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 1月17日

出 願 番 号

Application Number:

特願2001-009564

出 願 人

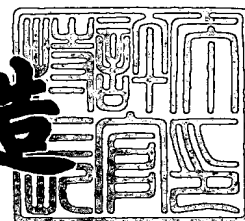
Applicant(s):

日本電気株式会社

2001年 9月25日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3087899

【書類名】 特許願

【整理番号】 68501900

【提出日】 平成13年 1月17日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 7/00

【発明者】

    【住所又は居所】 東京都港区芝五丁目7番1号 日本電気株式会社内

    【氏名】 長谷川 聡

【特許出願人】

    【識別番号】 000004237

    【氏名又は名称】 日本電気株式会社

【代理人】

    【識別番号】 100064621

    【弁理士】

    【氏名又は名称】 山川 政樹

    【電話番号】 03-3580-0961

【手数料の表示】

    【予納台帳番号】 006194

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

    【物件名】 図面 1

    【物件名】 要約書 1

    【包括委任状番号】 9718363

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 オーディオ情報処理装置及び処理方法

【特許請求の範囲】

【請求項 1】 入力した音声信号を含むオーディオ情報を複数の周波数帯域に分割するサブバンド分析部と、サブバンド分析部により各周波数帯域に分割された各サブバンドに対して基準値からの倍率を示すスケールファクタを計算し、ダイナミックレンジを揃えるスケーリング部と、スケーリング部の出力信号を M P E G 方式で圧縮符号化して符号化ビットストリームデータとして出力する符号化処理部とを備えたオーディオ情報処理装置において、

前記スケーリング部の出力であるスケールファクタ値に基づき前記オーディオ情報の特徴を抽出する特徴検出処理部を備えたことを特徴とするオーディオ情報処理装置。

【請求項 2】 請求項 1 において、

前記特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が音声信号区間か否かを判断する手段を備えたことを特徴とするオーディオ情報処理装置。

【請求項 3】 請求項 1 において、

前記特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が無音信号区間か否かを判断する手段を備えたことを特徴とするオーディオ情報処理装置。

【請求項 4】 請求項 1 において、

前記スケーリング部から出力される各サブバンド毎のスケールファクタ値を入力してこのスケールファクタ値に応じた信号レベルを算出する信号レベル算出部を備え、前記特徴検出処理部は、前記信号レベル算出部により算出された信号レベルに基づき前記オーディオ情報の特徴を抽出することを特徴とするオーディオ情報処理装置。

【請求項 5】 請求項 4 において、

前記信号レベル算出部は、前記スケーリング部から所定時間内に出力される低周波数帯域のスケールファクタ値を入力して信号レベルを算出し、

前記特徴検出処理部は、

前記信号レベル算出部により算出された信号レベルのうち最大値と最小値とを求めてこの最大値と最小値との差分を演算する演算手段と、

前記演算手段により演算された差分値が予め定めたとしきい値以上の場合は前記オーディオ情報を音声信号区間と判断し、前記差分値が前記しきい値未満の場合は前記オーディオ情報を音声以外の信号区間と判断する判断手段と

を有することを特徴とするオーディオ情報処理装置。

【請求項 6】 請求項 4 において、

前記信号レベル算出部は、前記スケーリング部から所定時間内に出力される全てのスケールファクタ値を入力して信号レベルを算出し、

前記特徴検出処理部は、前記信号レベル算出部により算出された信号レベルが予め定めたとしきい値以上の場合は有音信号区間と判断し、前記算出された信号レベルが前記しきい値未満の場合は無音信号区間と判断する判断手段を有することを特徴とするオーディオ情報処理装置。

【請求項 7】 M P E G 方式で符号化されたビットストリームデータを入力すると、各周波数帯域に分割された各サブバンドからなるこの符号化ビットストリームデータを各サブバンド毎に、ビット割当情報、基準値からの倍率を示すスケールファクタ値及び符号化データに分解するストリーム分解部と、ストリーム分解部により分解された符号化データを各サブバンド単位で復号処理しオーディオ情報として出力する復号化処理部とを備えたオーディオ情報処理装置において

前記ストリーム分解部の出力であるスケールファクタ値に基づき前記オーディオ情報の特徴を抽出する特徴検出処理部を備えたことを特徴とするオーディオ情報処理装置。

【請求項 8】 請求項 7 において、

前記特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が音声信号区間か否かを判断する手段を備えたことを特徴とするオーディオ情報処理装置。

【請求項 9】 請求項 7 において、

前記特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が無音信号区間か否かを判断する手段を備えたことを特徴とするオーディオ情報処理装置。

【請求項 1 0】 請求項 7 において、

前記ストリーム分解部から出力される各サブバンド毎のスケールファクタ値を入力して信号レベルを算出する信号レベル算出部を備え、前記特徴検出処理部は、前記信号レベル算出部により算出された信号レベルに基づき前記オーディオ情報の特徴を抽出することを特徴とするオーディオ情報処理装置。

【請求項 1 1】 請求項 1 0 において、

前記信号レベル算出部は、前記ストリーム分解部から所定時間内に出力される低周波数帯域のスケールファクタ値を入力して信号レベルを算出し、

前記特徴検出処理部は、

前記信号レベル算出部により算出された信号レベルのうち最大値と最小値とを求めてこの最大値と最小値との差分を演算する演算手段と、

前記演算手段により演算された差分値が予め定めたしきい値以上の場合は前記オーディオ情報を音声信号区間と判断し、前記差分値が前記しきい値未満の場合は前記オーディオ情報を音声以外の信号区間と判断する判断手段と

を有することを特徴とするオーディオ情報処理装置。

【請求項 1 2】 請求項 1 0 において、

前記信号レベル算出部は、前記ストリーム分解部から所定時間内に出力される全てのスケールファクタ値を入力して信号レベルを算出し、

前記特徴検出処理部は、前記信号レベル算出部により算出された信号レベルが予め定めたしきい値以上の場合は有音信号区間と判断し、前記算出された信号レベルが前記しきい値未満の場合は無音信号区間と判断する判断手段を有することを特徴とするオーディオ情報処理装置。

【請求項 1 3】 入力した音声信号を含むオーディオ情報を複数の周波数帯域に分割するサブバンド分析部と、サブバンド分析部により各周波数帯域に分割された各サブバンドに対して基準値からの倍率を示すスケールファクタを計算し、ダイナミックレンジを揃えるスケーリング部と、スケーリング部の出力信号を

MPEG方式で圧縮符号化して符号化ビットストリームデータとして出力する符号化処理部とを備えたオーディオ情報処理装置において、

前記スケーリング部の出力であるスケールファクタ値に基づき前記オーディオ情報の特徴を抽出する第1のステップを有することを特徴とする処理方法。

【請求項14】 請求項13において、

前記第1のステップにおける処理は、スケールファクタ値に基づき前記オーディオ情報が音声信号区間か否かを判断する第2のステップを含むことを特徴とする処理方法。

【請求項15】 請求項13において、

前記第1のステップにおける処理は、スケールファクタ値に基づき前記オーディオ情報が無音信号区間か否かを判断する第3のステップを含むことを特徴とする処理方法。

【請求項16】 請求項13において、

前記スケーリング部から出力される各サブバンド毎のスケールファクタ値を入力してこのスケールファクタ値に応じた信号レベルを算出する第4のステップを有し、

前記第1のステップにおける処理は、第4のステップの処理により算出された信号レベルに基づき前記オーディオ情報の特徴を抽出する第5のステップを含むことを特徴とする処理方法。

【請求項17】 請求項16において、

前記第4のステップにおける処理は、前記スケーリング部から所定時間内に出力される低周波数帯域のスケールファクタ値を入力して信号レベルを算出する第6のステップを含み、

前記第5のステップにおける処理は、

前記第6のステップの処理に基づき算出された信号レベルのうち最大値と最小値とを求めてこの最大値と最小値との差分を演算する第7のステップと、

前記第7のステップの処理に基づく差分値が予め定めたしきい値以上の場合は前記オーディオ情報を音声信号区間と判断し、前記差分値が前記しきい値未満の場合は前記オーディオ情報を音声以外の信号区間と判断する第8のステップと

を含むことを特徴とする処理方法。

【請求項 1 8】 請求項 1 6 において、

前記第 4 のステップにおける処理は、前記スケーリング部から所定時間内に出力される全てのスケールファクタ値を入力して信号レベルを算出する第 9 のステップを含み、

前記第 5 のステップにおける処理は、前記第 9 のステップの処理に基づき算出された信号レベルが予め定めたしきい値以上の場合は有音信号区間と判断し、前記算出された信号レベルが前記しきい値未満の場合は無音信号区間と判断する第 1 0 のステップを含むことを特徴とする処理方法。

【請求項 1 9】 M P E G 方式で符号化されたビットストリームデータを入力すると、各周波数帯域に分割された各サブバンドからなるこの符号化ビットストリームデータを各サブバンド毎に、ビット割当情報、基準値からの倍率を示すスケールファクタ値及び符号化データに分解するストリーム分解部と、ストリーム分解部により分解された符号化データを各サブバンド単位で復号処理しオーディオ情報として出力する復号化処理部とを備えたオーディオ情報処理装置において、

前記ストリーム分解部の出力であるスケールファクタ値に基づき前記オーディオ情報の特徴を抽出する第 1 1 のステップを有することを特徴とする処理方法。

【請求項 2 0】 請求項 1 9 において、

前記第 1 1 のステップにおける処理は、スケールファクタ値に基づき前記オーディオ情報が音声信号区間か否かを判断する第 1 2 のステップを含むことを特徴とする処理方法。

【請求項 2 1】 請求項 1 9 において、

前記第 1 1 のステップにおける処理は、スケールファクタ値に基づき前記オーディオ情報が無音信号区間か否かを判断する第 1 3 のステップを含むことを特徴とする処理方法。

【請求項 2 2】 請求項 1 9 において、

前記ストリーム分解部から出力される各サブバンド毎のスケールファクタ値を入力して信号レベルを算出する第 1 4 のステップを有し、



前記第 1 1 のステップにおける処理は、前記第 1 4 のステップの処理により算出された信号レベルに基づき前記オーディオ情報の特徴を抽出する第 1 5 のステップを含むことを特徴とする処理方法。

【請求項 2 3】 請求項 2 2 において、

前記第 1 4 のステップにおける処理は、前記ストリーム分解部から所定時間内に出力される低周波数帯域のスケールファクタ値を入力して信号レベルを算出する第 1 6 のステップを含み、

前記第 1 5 のステップにおける処理は、

前記第 1 6 のステップの処理に基づき算出された信号レベルのうち最大値と最小値とを求めてこの最大値と最小値との差分を演算する第 1 7 のステップと、

前記第 1 7 のステップの処理に基づく差分値が予め定めたしきい値以上の場合は前記オーディオ情報を音声信号区間と判断し、前記差分値が前記しきい値未満の場合は前記オーディオ情報を音声以外の信号区間と判断する第 1 8 のステップと

を含むことを特徴とする処理方法。

【請求項 2 4】 請求項 2 2 において、

前記第 1 4 のステップにおける処理は、前記ストリーム分解部から所定時間内に出力される全てのスケールファクタ値を入力して信号レベルを算出する第 1 9 のステップを含み、

前記第 1 5 のステップにおける処理は、前記第 1 9 のステップの処理に基づき算出された信号レベルが予め定めたしきい値以上の場合は有音信号区間と判断し、前記算出された信号レベルが前記しきい値未満の場合は無音信号区間と判断する第 2 0 のステップを含むことを特徴とする処理方法。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、オーディオ情報処理装置及び処理方法に関する。

【0 0 0 2】

【従来の技術】

近年、パーソナルコンピュータ等の性能向上やインターネットの普及により、マルチメディア情報を広く利用することが可能になってきている。これと同時に、マルチメディア情報の効率の良い検索や、所望の情報を抽出するなどの要求が増えてきており、重要なテーマとなっている。特に、映像情報や音声情報に対する要求は、いわゆるデジタル家電と呼ばれるデジタルビデオカメラやデジタルスチルカメラのような製品の普及に伴って急速に増加しており、今後もその需要の増加が期待される。

## 【0003】

音声情報に関する情報検索手法や情報抽出手法は、MPEG (Moving Picture Experts Group) 方式等で圧縮符号化されたオーディオ情報に対するものや、符号化されていないオーディオ情報に対するものなど、多くの方式が提案されている。

## 【0004】

例えば特開平10-247093号公報では、符号化されていないオーディオ情報及びMPEG方式による圧縮符号化されたオーディオ情報の双方に対し、そのオーディオ情報を音楽区間と音声信号区間に分類するオーディオ情報分類装置が提案されている。この装置によれば、符号化されていないオーディオ情報からは、単位時間毎の周波数データを抽出し、これから単位時間当たりのエネルギーを求め、その区間が音声信号区間なのか音楽区間なのかを判断している。一方、MPEG方式で圧縮符号化されているオーディオ情報からは、各フレームのサブバンドデータを復号し、このサブバンドデータから単位時間当たりのエネルギーを求め、その区間が音声信号区間なのか音楽区間なのかを判断している。

## 【0005】

また、特開2000-66691号公報では、符号化されていないオーディオ情報及びMPEG方式による圧縮符号化されたオーディオ情報の双方に対し、そのオーディオ情報を音声信号区間、音楽区間、雑音区間に分類するオーディオ情報分類装置が提案されている。この装置は、特開平10-247093号公報の装置と同様の処理で単位時間当たりのエネルギーを求めた後、そのエネルギーの分散や疎密度、及び重心から音声信号区間、音楽区間、雑音区間を判断している。

## 【0006】

図6は、MPEG1/Audioレイヤ1方式（ISO/IEC 11172-3）を適用した符号化処理装置の構成を示すブロック図であり、サブバンド分析部111と、スケーリング部112と、ビット割当部113と、量子化部114と、ビットストリーム生成部115と、聴覚心理モデル（心理聴覚分析部）116とから構成される。ここで、サブバンド分析部111は入力信号Aを複数の周波数帯域に分割する一方、スケーリング部112は分割された各サブバンド信号に対して基準値からの倍率を示すスケールファクタを計算してダイナミックレンジを揃えけるとともに、聴覚心理モデル116は各サブバンドで音声信号がマスキングされている比率を求め、ビット割当部113は聴覚心理モデル116の出力結果をもとに各サブバンドへのビット割当を行う。そして、量子化部114はビット割当部3の出力を量子化計算し、ビットストリーム生成部115は量子化部114からの出力情報にヘッダや補助情報を付加して音声符号化データBとして出力するようにしている。

## 【0007】

## 【発明が解決しようとする課題】

しかしながら、前述した図6に示すようなMPEG方式を適用したオーディオ情報の圧縮符号化処理を行う符号化処理装置では、オーディオ情報の符号化処理中に、その入力オーディオ情報の音声信号区間や無音信号区間等の特徴抽出が行えないという課題があった。

したがって、本発明は、入力したオーディオ情報の符号化処理中に、その入力オーディオ情報の特徴抽出を可能にすることを目的とする。

## 【0008】

## 【課題を解決するための手段】

このような課題を解決するために本発明は、入力した音声信号を含むオーディオ情報を複数の周波数帯域に分割するサブバンド分析部と、サブバンド分析部により各周波数帯域に分割された各サブバンドに対して基準値からの倍率を示すスケールファクタを計算し、ダイナミックレンジを揃えるスケーリング部と、スケ

ーリング部の出力信号を圧縮符号化して符号化ビットストリームデータとして出力する符号化処理部とを備えたオーディオ情報処理装置において、スケーリング部から出力されるスケールファクタ値をもとにオーディオ情報の特徴を抽出する特徴検出処理部を設けたものである。

この場合、特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が音声信号区間か否かを判断するものである。

また、特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が無音信号区間か否かを判断するものである。

また、スケーリング部から出力される各サブバンド毎のスケールファクタ値を入力してこのスケールファクタ値に応じた信号レベルを算出する信号レベル算出部を設け、特徴検出処理部は、信号レベル算出部により算出された信号レベルに基づきオーディオ情報の特徴を抽出するものである。

#### 【0009】

また、信号レベル算出部は、スケーリング部から所定時間内に出力される低周波数帯域のスケールファクタ値を入力して信号レベルを算出し、特徴検出処理部は、信号レベル算出部により算出された信号レベルのうち最大値と最小値とを求めてこの最大値と最小値との差分を演算する演算手段と、演算手段により演算された差分値が予め定めたしきい値以上の場合はオーディオ情報を音声信号区間と判断し、差分値がしきい値未満の場合はオーディオ情報を音声以外の信号区間と判断する判断手段とを有するものである。

また、信号レベル算出部は、スケーリング部から所定時間内に出力される全てのスケールファクタ値を入力して信号レベルを算出し、特徴検出処理部は、信号レベル算出部により算出された信号レベルが予め定めたしきい値以上の場合は有音信号区間と判断し、算出された信号レベルがしきい値未満の場合は無音信号区間と判断する判断手段を有するものである。

#### 【0010】

また、本発明は、符号化ビットストリームデータを入力すると、各周波数帯域に分割された各サブバンドからなるこの符号化ビットストリームデータを各サブバンド毎に、ビット割当情報、基準値からの倍率を示すスケールファクタ値及び

符号化データに分解するストリーム分解部と、ストリーム分解部により分解された符号化データを各サブバンド単位で復号処理しオーディオ情報として出力する復号化処理部とを備えたオーディオ情報処理装置において、ストリーム分解部から出力されるスケールファクタ値をもとにオーディオ情報の特徴を抽出する特徴検出処理部を設けたものである。

この場合、特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が音声信号区間か否かを判断するものである。

また、特徴検出処理部は、スケールファクタ値に基づき前記オーディオ情報が無音信号区間か否かを判断するものである。

また、ストリーム分解部から出力される各サブバンド毎のスケールファクタ値を入力して信号レベルを算出する信号レベル算出部を設け、特徴検出処理部は、信号レベル算出部により算出された信号レベルに基づきオーディオ情報の特徴を抽出するものである。

#### 【 0 0 1 1 】

また、信号レベル算出部は、ストリーム部から所定時間内に出力される低周波数帯域のスケールファクタ値を入力して信号レベルを算出し、特徴検出処理部は、信号レベル算出部により算出された信号レベルのうち最大値と最小値とを求めてこの最大値と最小値との差分を演算する演算手段と、演算手段により演算された差分値が予め定めたしきい値以上の場合はオーディオ情報を音声信号区間と判断し、差分値がしきい値未満の場合はオーディオ情報を音声以外の信号区間と判断する判断手段とを有するものである。

また、信号レベル算出部は、ストリーム分解部から所定時間内に出力される全てのスケールファクタ値を入力して信号レベルを算出し、特徴検出処理部は、信号レベル算出部により算出された信号レベルが予め定めたしきい値以上の場合は有音信号区間と判断し、算出された信号レベルがしきい値未満の場合は無音信号区間と判断する判断手段を有するものである。

#### 【 0 0 1 2 】

#### 【発明の実施の形態】

以下、本発明について図面を参照して説明する。

図1は、本発明を適用した符号化処理装置の構成を示すブロック図であり、MPEG (Moving Picture Experts Group) 1/Audioレイヤ1方式 (ISO/IEC 11172-3) の符号化処理装置の構成を示すものである。この符号化処理装置は、図1に示すように、入力信号 (入力音声データ) a を複数の周波数帯域に分割するサブバンド分析部11と、異なる周波数帯域に分割された各サブバンド信号に対して基準値からの倍率を示すスケールファクタを計算し、ダイナミックレンジを揃えるスケーリング部12と、入力信号aとスケーリング部12の出力とに基づき各サブバンドで音声信号がマスキングされている比率を求める聴覚心理モデル (心理聴覚分析部) 16と、聴覚心理モデル16からの出力結果に基づき、スケーリング部12からの各サブバンドに対するビット割り当てを行うビット割当部13と、ビット割当部13の出力を量子化計算する量子化部14と、量子化部14により量子化されたデータにヘッダや補助情報を付加してビット列を形成し音声符号化データbとして出力するビットストリーム生成部15と、スケーリング部12で得られたスケールファクタ値をもとにして音声情報を抽出するサウンド情報抽出部20とからなる。

## 【0013】

ここで、サウンド情報抽出部20は、スケールファクタ値から各サブバンド単位での信号レベルを計算する信号レベル算出部21と、信号レベル算出部21により算出された信号レベルに基づき、入力信号aに対する無音検出や音声検出等の分析処理を行う特徴検出処理部22とから構成される。

## 【0014】

特徴検出処理部22では、定められた時間範囲における信号レベルの最大値と最小値の差分をしきい値で比較する第1の方法、定められた時間範囲における信号レベルが全てしきい値を下回っているかどうかを判断する第2の方法、求められた信号レベルの絶対値としきい値とを比較する第3の方法、過去の信号レベル履歴からその絶対値や振幅等の平均値もしくは分散を求め得られた結果の変化量を判断する第4の方法、及び各サブバンド毎の信号レベル比率を比較する第5の方法などにより入力信号の分析処理を行い入力信号aの特徴を抽出する。

## 【0015】

## (第1の実施の形態)

次に、図1のブロック図及び図3のフローチャートに基づき本発明の第1の実施の形態を説明する。第1の実施の形態では、MPEG1/Audioレイヤ1の場合を例に、音声検出の場合について説明する。

16ビットの直線量子化された入力信号aは、図1のサブバンド分析部11で32帯域のサブバンド信号に分割される。各サブバンド当たり12サンプルが抽出され、合計384(32×12)サンプル単位で以降の処理が実行される。この32帯域に分割された各サブバンド信号のダイナミックレンジを揃えるため、図1のスケーリング部12では最大振幅(基準値)が1.0になるように正規化し、その倍率を示すスケールファクタ値を各サブバンド単位で算出する。

## 【0016】

ここで、算出されるスケールファクタ値とは、最大振幅1.0に対する実際の信号との割合を示したものであるが、この値が大きいほど大きな振幅を持った信号であることが各サブバンド単位で判断できることになる。スケーリング部12で得られたスケールファクタ値は、図1の心理聴覚モデル16とビット割当部13に渡され、符号化処理が継続される一方、サウンド情報抽出部20にも渡されサウンド情報(音声情報、オーディオ情報)の抽出処理に使用される。

## 【0017】

この場合、サウンド情報抽出部20の信号レベル算出部21は、スケーリング部12から図3のステップS1で各サブバンド毎のスケールファクタ値を取得する。そして、取得したスケールファクタ値のうち低周波数帯域側のスケールファクタ値を得て低周波数帯域の信号レベルを算出する(ステップS2)。これは、音声信号の周波数帯域が狭く、低周波数帯域に集中していることによる。信号レベル算出方法の一例としては、MPEG/Audioの規格書であるISO/IEC11172-3に、各サブバンド当たりの信号レベル計算式が提案されている。

## 【0018】

即ち、各サブバンド当たりの音圧レベルをLsbとすると、

$$Lsb(n) = 20 \times \log(Scfmax(n) \times 32768)$$

- 10

. . . . . (1)

となる信号レベルの計算式が提案されている。ここで、 $n$ はサブバンド番号、 $S_{cfmax}(n)$ は各サブバンド毎のスケールファクタ値である。本実施の形態では式(1)を用いているが、信号レベル算出式はこの計算式に限るものではない。

## 【0019】

こうして信号レベル算出部21により算出された低周波数帯域の信号レベルは、特徴検出処理部22に渡される。特徴検出処理部22では、まず今回取得した信号レベルが今まで取得した信号レベルのなかで最大値もしくは最小値であるかの判断を行う(ステップS3)。最大値もしくは最小値と判断されステップS3で「Y」となる場合は、新たな最大値もしくは最小値としてその信号レベルを格納する(ステップS4)。取得した信号レベルが今まで取得した信号レベルのなかで最大値もしくは最小値ではなくステップS3の判定が「N」となる場合は、今回取得した信号レベルは格納しない。

## 【0020】

次に、ステップS5では1秒分の信号レベルを確認したか否かの判断を行う。本実施の形態では1秒単位での音声検出を実施することとする。なお、MPEG/Audioレイヤ1の場合、サンプリング周波数が44.1kHzであれば384サンプル当たり約8.7ミリ秒である。1秒分の信号レベルを確認したと判断し、ステップS5の判定が「Y」となると、現在までに格納されている信号レベルの最大値と最小値の差分を求める(ステップS6)。

## 【0021】

そして、信号レベルの最大値と最小値の差分値が予め定められたしきい値以上でありステップS7の判定が「Y」となると、特徴検出処理部22はこの1秒間は音声信号区間であると判断し、音声信号区間としてのパラメータcを出力する(ステップS8)。一方、信号レベルの最大値と最小値の差分値が予め定められたしきい値未満であり、ステップS7の判定が「N」となる場合は、特徴検出処理部22はこの1秒間は例えば音楽等の音声以外の信号区間であると判断し、音声以外の信号区間としてのパラメータcを出力する(ステップS9)。そして、



パラメータが出力された後、現在格納されている信号レベルの最大値と最小値がリセットされ（ステップ S 1 0）、次の 1 秒間についての検出が開始される。なお、まだ 1 秒分の信号レベルを確認していないと判断されステップ S 5 の判定が「N」となる場合は、次の 3 8 4 サンプル当たりのスケールファクタ値が取得され同様に処理される。

## 【 0 0 2 2 】

（第 2 の実施の形態）

次に、図 1 のブロック図及び図 4 のフローチャートを用いて本発明の第 2 の実施の形態について説明する。

第 1 の実施の形態では、サウンド情報抽出部 2 0 の処理により音声信号区間を検出する場合の例を説明したが、第 2 の実施の形態では、このサウンド情報抽出部 2 0 の処理により無音信号区間を検出する場合について説明する。

1 6 ビットの直線量子化された入力信号 a は、図 1 のサブバンド分析部 1 1 で同様に 3 2 帯域のサブバンド信号に分割される。この 3 2 帯域に分割された各サブバンド信号のダイナミックレンジを揃えるため、図 1 のスケーリング部 1 2 では最大振幅が 1. 0 になるように正規化し、その倍率を示すスケールファクタ値を各サブバンド単位で算出する。そして、スケーリング部 1 2 で得られたスケールファクタ値は、同様に図 1 の心理聴覚モデル 1 6 とビット割当部 1 3 に渡されて符号化処理が継続される一方、サウンド情報抽出部 2 0 にも渡されサウンド情報の抽出処理に使用される。

## 【 0 0 2 3 】

この場合、サウンド情報抽出部 2 0 の信号レベル算出部 2 1 では、スケーリング部 1 2 から図 4 のステップ S 1 1 で各サブバンド毎のスケールファクタ値を取得する。そして、取得した全てのスケールファクタ値を用い 3 8 4 サンプル当たりの信号レベルを算出する（ステップ S 1 2）。ここで、信号レベルの算出の際の算出式については、前述の式（1）を用いても良いし、この算出式に限るものではない。

## 【 0 0 2 4 】

次に、特徴検出処理部 2 2 では、信号レベル算出部 2 1 で求められた 3 8 4 サ

ンプル当たりの信号レベルが予め定められたしきい値未満であるか否かの判断を行う（ステップ S 1 3）。ここで、前記信号レベルが予め定められたしきい値未満ではないと判断され、ステップ S 1 3 の判定が「N」となる場合は、特徴検出処理部 2 2 は有音信号区間と判断して有音信号区間としてのパラメータ c を出力する（ステップ S 1 4）。そしてその後、次の 3 8 4 サンプル当たりのスケールファクタ値が取得され処理される。

## 【 0 0 2 5 】

一方、信号レベル算出部 2 1 で求められた 3 8 4 サンプル当たりの信号レベルがしきい値未満であると判断されステップ S 1 3 の判定が「Y」となる場合は、そのしきい値未満の状態が 1 秒間以上継続しているか否かを判断する（ステップ S 1 5）。ここで、1 秒間以上継続していると判断されステップ S 1 5 で「Y」となる場合は、特徴検出処理部 2 2 はその区間を無音信号区間と判断して、無音信号区間としてのパラメータ c を出力する（ステップ S 1 6）。そしてその後、次の 3 8 4 サンプル当たりのスケールファクタ値が取得され同様に処理される。なお、しきい値未満の信号レベルが 1 秒以上継続していないと判断されステップ S 1 5 で「N」となる場合は次の 3 8 4 サンプル当たりのスケールファクタ値が取得され処理される。

## 【 0 0 2 6 】

このように、第 1 及び第 2 の実施の形態では、M P E G 方式の音声符号化処理で算出されたパラメータのスケールファクタ値を使用してサウンド情報（音声情報、オーディオ情報）特徴抽出処理を行うため、サウンド情報抽出処理で用いられる専用の特別なパラメータの抽出処理が不要になり、したがって軽負荷で処理できる。このため、実時間で音声符号化処理するような場合であっても、同時にサウンド情報を抽出できる。

## 【 0 0 2 7 】

## （第 3 の実施の形態）

図 2 は、本発明を適用した復号化装置のブロック図であり、M P E G 符号化されたデータの特徴抽出を行うための構成を示すものである。図 2 において、本復号化装置は、入力符号化データ b を各サブバンド当たりのビット割当情報、スケ

ールファクタ値、及び符号化データに分解するビットストリーム分解部（以下、ストリーム分解部）31と、ストリーム分解部31により分解されたデータを各サブバンド単位で復号処理する逆量子化部32と、逆量子化部32により復号化された各サブバンドを合成し音声データdとして出力するサブバンド合成部33と、ストリーム分解部31に分解されたスケールファクタ値をもとにして音声情報を抽出する、前述の信号レベル算出部21及び特徴検出処理部22からなる前記サウンド情報抽出部20とから構成される。

## 【0028】

次に、図2のブロック図、図3、図4のフローチャート及び図5のデータフォーマットを用いて本発明の第3の実施の形態について説明する。前述の第1及び第2の実施の形態では、サウンド情報抽出部20による処理を符号化処理中に組みこんだが、第3の実施の形態ではMPEG符号化されたオーディオビットストリームからサウンド情報抽出部20によりサウンド情報を抽出する場合について説明する。

## 【0029】

まず、MPEG方式による音声復号処理手法について説明する。この実施の形態では、MPEG/Audioレイヤ1の場合を例に説明する。MPEG符号化されたビットストリームは、図5に示すように、先頭から順に、ヘッダ41、エラーチェック情報42、ビット割当情報43、スケールファクタ値44及び符号化データ45が割り当てられるようなデータフォーマットを有している。このようなビットストリームを受信すると、まず図2のストリーム分解部31において各サブバンド当たりのビット割当情報、スケールファクタ値、及び符号化データに分解される。そして、逆量子化部32において各サブバンド単位での復号処理が行われた後、サブバンド合成部33は各サブバンドを合成し、音声信号として出力する。

## 【0030】

ところで、従来は、サウンド情報を抽出するためには、サブバンド合成部33から出力される音声信号を使用するか、もしくは逆量子化部32で各サブバンド単位で復号された情報を使用するかの方法が採られていたが、本実施の形態では

、まずストリーム分解部 3 1 でビットストリームを分解し、ここで得られたスケールファクタ値をそのままサウンド情報抽出部 2 0 に渡す。以降、サウンド情報抽出部 2 0 の信号レベル算出部 2 1 と特徴検出処理部 2 2 でサウンド情報抽出処理を行うが、その抽出処理は前述した第 1 及び第 2 の実施の形態と同様の処理である。

#### 【 0 0 3 1 】

即ち、信号レベル算出部 2 1 は、ストリーム分解部 3 1 からの低周波数帯域のスケールファクタ値を入力すると信号レベルを算出し、特徴検出処理部 2 2 は、信号レベル算出部 2 1 により算出された信号レベルのうち最大値と最小値とを求めてこの最大値と最小値との差分を演算し、その差分値が予め定めたしきい値以上の場合は符号化データ b を音声信号と判断し、差分値がしきい値未満の場合は符号化データ b を音声以外の信号と判断する（第 1 の実施の形態に対応する処理）。

また、信号レベル算出部 2 1 は、ストリーム分解部 3 1 からの全てのスケールファクタ値を入力してその信号レベルを算出し、特徴検出処理部 2 2 は、信号レベル算出部 2 1 により算出された信号レベルが予め定めたしきい値以上の場合は有音信号と判断し、算出された信号レベルがしきい値未満の場合は無音信号と判断する（第 2 の実施の形態に対応する処理）。

#### 【 0 0 3 2 】

このように、MPEG 方式で圧縮符号化されたビットストリームからサウンド情報を抽出する場合、ビットストリーム中のスケールファクタ値をパラメータとして使用するようにしたので、ビットストリームの復号処理を行うことなくサウンド情報の抽出処理を行うことができる。したがって、軽負荷でサウンド情報を抽出できることから、性能の低いパーソナルコンピュータでも高速な処理が可能である。

#### 【 0 0 3 3 】

以上説明したように、本発明は、オーディオ情報圧縮符号化処理中に処理負荷を軽減した効率の良いオーディオ情報抽出方法を提供できるとともに、MPEG 方式で符号化されたデータであっても、ビットストリーム解析処理のみでオーデ

ィオ情報の抽出を可能にすることができる。

なお、前述した第1～第3の実施の形態では、MPEG/Audioレイヤ1の場合について述べたが、本発明はサブバンド分割による符号化方式でスケールファクタ値を算出する手段を有する他の音声符号化方式、例えばMPEG/Audioレイヤ2、MPEG/Audioレイヤ3などにも適用可能である。

#### 【0034】

また、サウンド情報抽出部20内の特徴検出処理部22では、第1の実施の形態で説明したような、定められた時間範囲における信号レベルの最大値と最小値の差分をしきい値で比較する方法や、第2の実施の形態で説明したような、定められた時間範囲における信号レベルが全てしきい値を下回っているかどうかを判断する方法の他に、信号レベルの絶対値としきい値とを比較する方法や、過去の信号レベル履歴からその絶対値や振幅等の平均値もしくは分散を求め、得られた結果の変化量から判断する方法などを用いて、入力信号の分析処理を行ってもよい。

#### 【0035】

##### 【発明の効果】

以上説明したように本発明は、入力した音声信号を含むオーディオ情報を複数の周波数帯域に分割するサブバンド分析部と、サブバンド分析部により各周波数帯域に分割された各サブバンドに対して基準値からの倍率を示すスケールファクタを計算し、ダイナミックレンジを揃えるスケーリング部と、スケーリング部の出力信号を圧縮符号化して符号化ビットストリームデータとして出力する符号化処理部とを備えたオーディオ情報処理装置において、スケーリング部から出力される各サブバンド毎のスケールファクタ値に基づきオーディオ情報の特徴を抽出するようにしたので、入力したオーディオ情報の符号化処理中に、同時にそのオーディオ情報の特徴抽出が可能になる。また、本発明ではMPEG方式の音声符号化処理で算出されたパラメータのスケールファクタ値を使用してサウンド情報（オーディオ情報）特徴抽出処理を行うため、サウンド情報抽出処理で用いられる専用の特別なパラメータを抽出するための処理が不要になり、したがって軽負荷で処理できる。このため、実時間で音声符号化処理するような場合であっても

、同時にサウンド情報を抽出できる。

【0036】

また、符号化ビットストリームデータを入力すると、各周波数帯域に分割された各サブバンドからなるこの符号化ビットストリームデータを各サブバンド毎に、ビット割当情報、基準値からの倍率を示すスケールファクタ値及び符号化データに分解するストリーム分解部と、ストリーム分解部により分解された符号化データを各サブバンド単位で復号処理しオーディオ情報として出力する復号化処理部とを備えたオーディオ情報処理装置において、ストリーム分解部から出力される各サブバンド毎のスケールファクタ値に基づきオーディオ情報の特徴を抽出するようにしたので、符号化されたビットストリームを復号化することなくサウンド情報を抽出できる。したがって、軽負荷でサウンド情報を抽出できることから、性能の低いパーソナルコンピュータでも高速処理が期待できる。

【図面の簡単な説明】

【図1】 本発明を適用したMPEG/Audioレイヤ1符号化方式の符号化装置の構成を示すブロック図である。

【図2】 MPEG/Audioレイヤ1符号化方式で符号化されたオーディオビットストリームを復号化する復号化装置の構成を示すブロック図である。

【図3】 図1及び図2に示す装置による音声検出を行う場合の動作を示すフローチャートである。

【図4】 図1及び図2に示す装置による無音検出を行う場合の動作を示すフローチャートである。

【図5】 MPEG/Audioレイヤ1符号化方式で符号化されたオーディオビットストリームのフォーマットを示す図である。

【図6】 MPEG/Audioレイヤ1符号化方式を適用した符号化装置の構成を示すブロック図である。

【符号の説明】

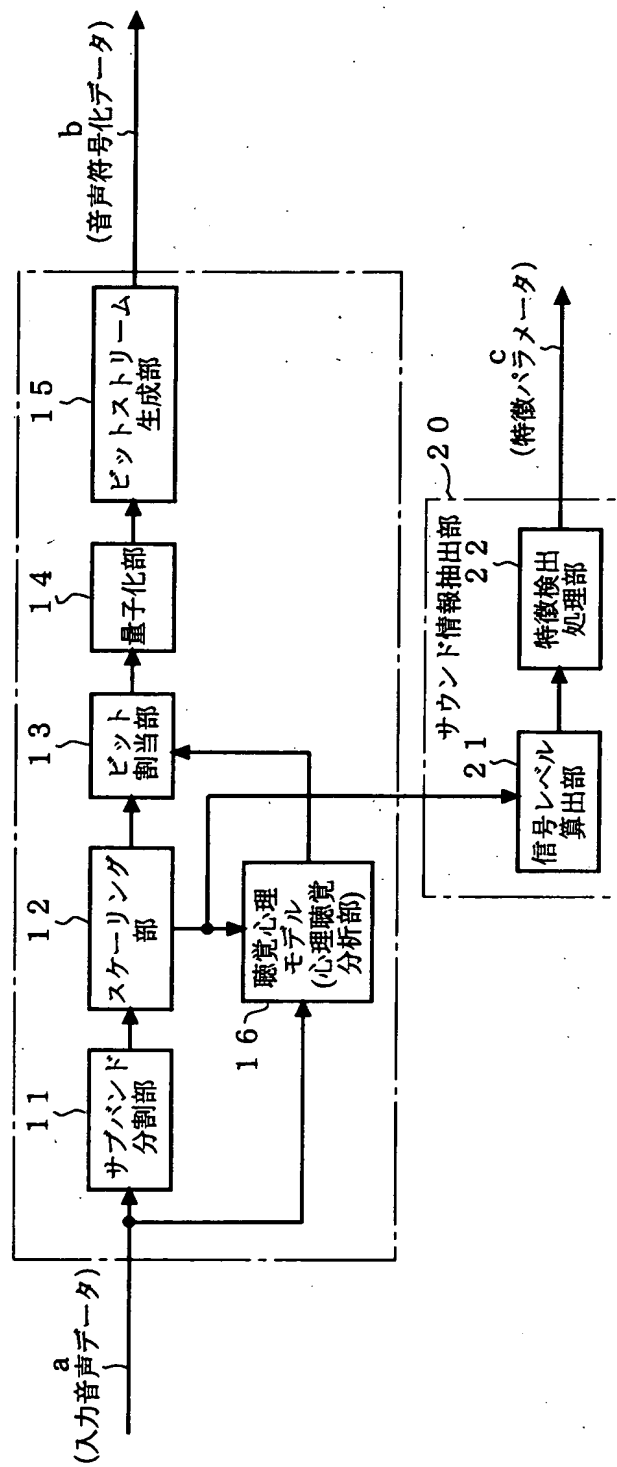
11…サブバンド分析部、12…スケーリング部、13…ビット割当部、14…量子化部、15…ビットストリーム生成部、16…心理聴覚モデル、20…サウンド情報抽出部、21…信号レベル算出部、22…特徴検出処理部、31…ピ

ットストリーム分解部、 3 2 …逆量子化部、 3 3 …サブバンド合成部。

【書類名】

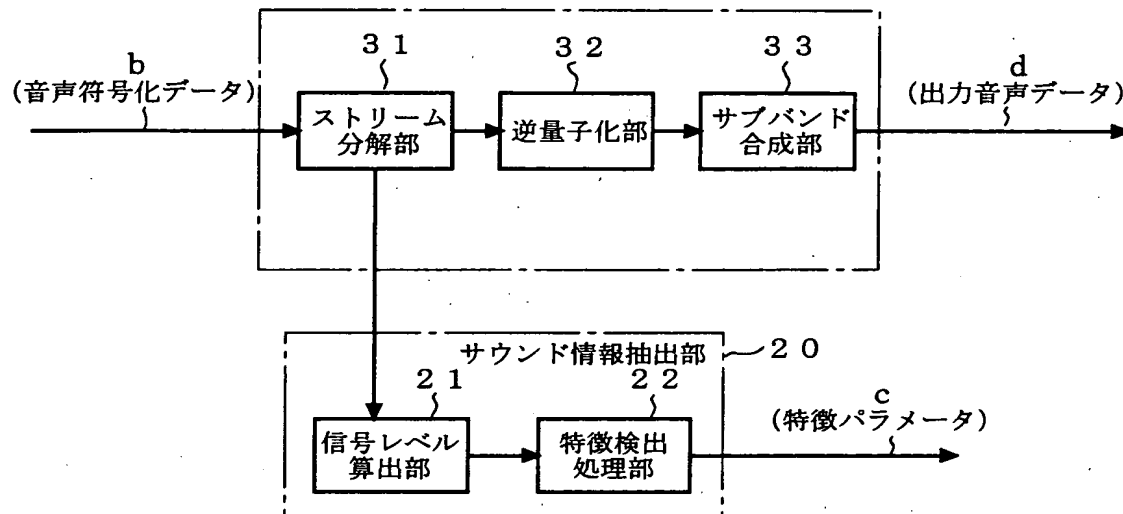
図面

【図 1】

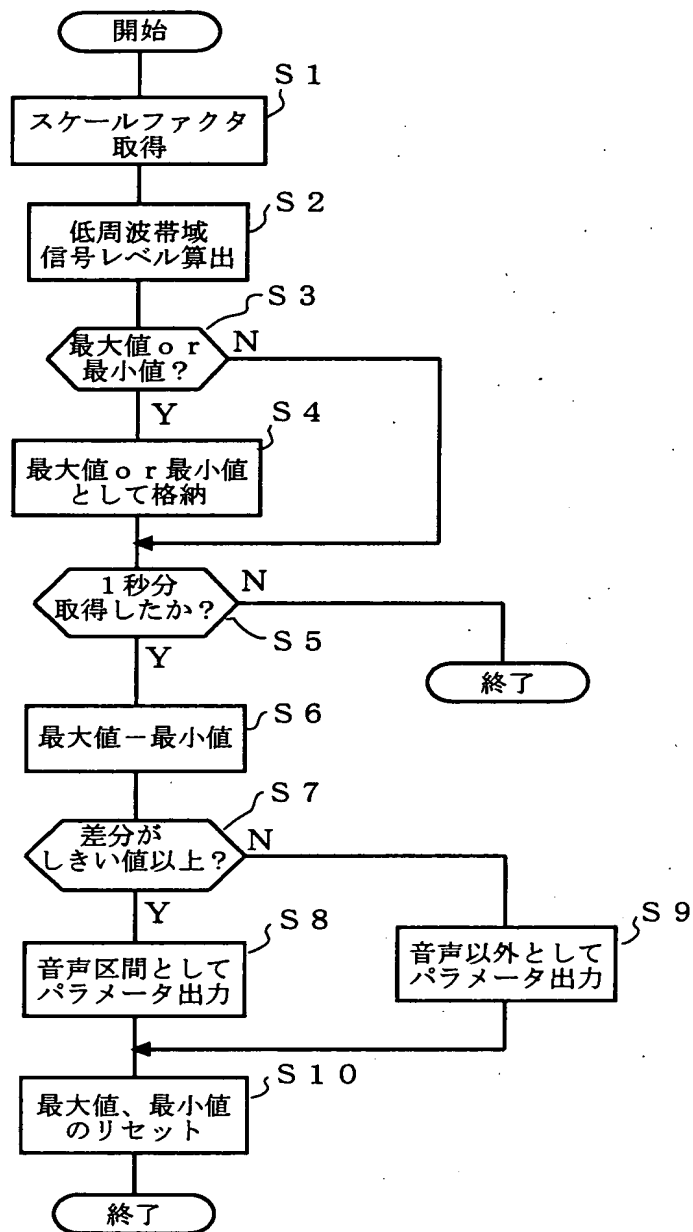




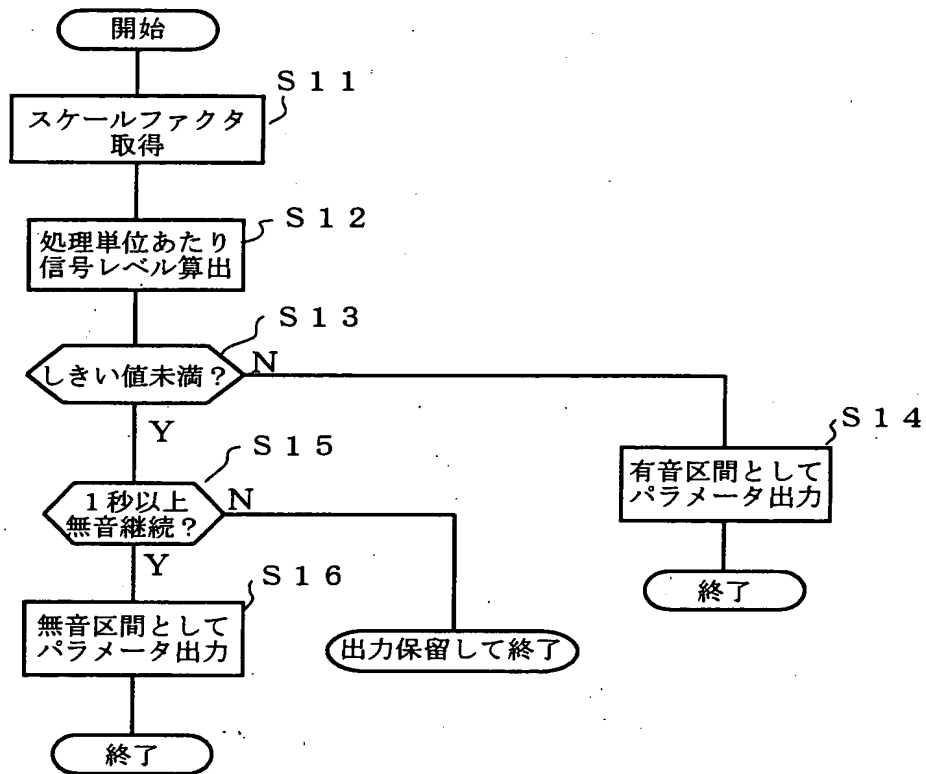
【図 2】



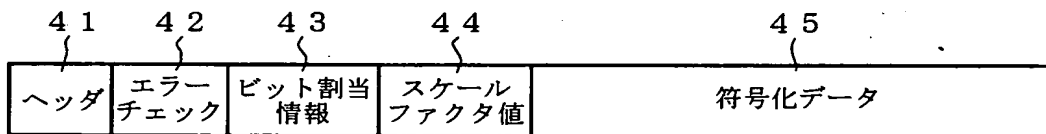
【図 3】



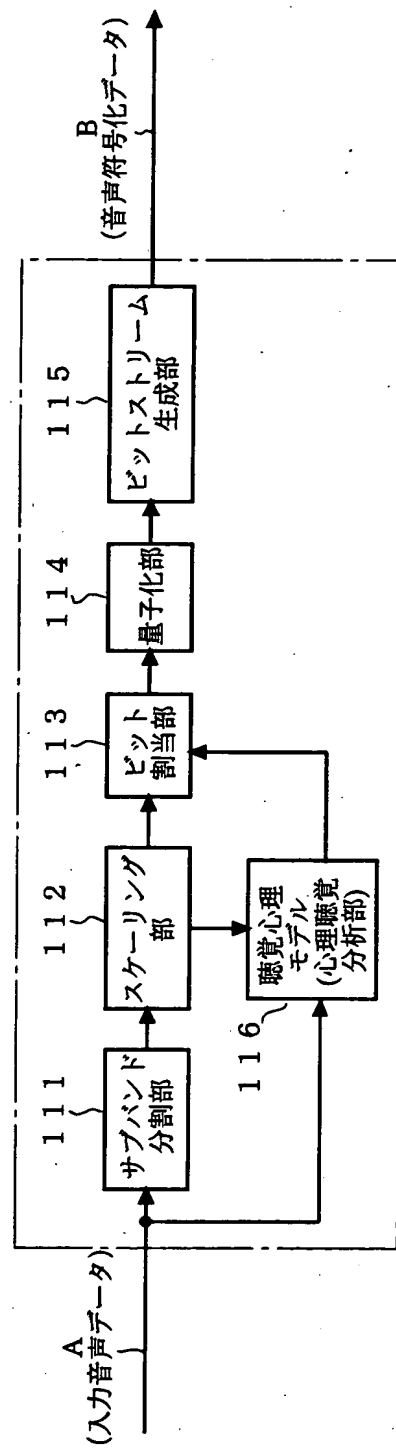
【図 4】



【図 5】



【図 6】



【書類名】            要約書

【要約】

【課題】    入力したオーディオ情報の符号化処理中に、その入力オーディオ情報の特徴抽出を可能にする。

【解決手段】    入力したオーディオ情報が複数の周波数帯域に分割された各サブバンド信号に対して、スケーリング部 1 2 により基準値からの倍率を示すスケールファクタを計算してダイナミックレンジを揃え、スケーリング部の出力信号を M P E G 方式で符号化する装置に、スケーリング部からの各サブバンド毎のスケールファクタ値を入力して信号レベルを算出する信号レベル算出部 2 1 と、算出された信号レベルのうち最大値、最小値を求めてこの最大値と最小値との差分を計算し、この差分値が予め定めたしきい値以上の場合に音声信号と判定し、しきい値未満の場合は音声以外の信号と判定する特徴検出処理部 2 2 とを設ける。

【選択図】            図 3

出 願 人 履 歴 情 報

識別番号 [000004237]

1. 変更年月日 1990年 8月29日

[変更理由] 新規登録

住 所 東京都港区芝五丁目7番1号

氏 名 日本電気株式会社